Available online at www.sciencedirect.com

ScienceDirect

Journal of Functional Analysis 258 (2010) 3758–3774

JOURNAL OF Functional Analysis

www.elsevier.com/locate/jfa

Duality on gradient estimates and Wasserstein controls

Kazumasa Kuwada [*,1,2]

Institut für Angewandte Mathematik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany

Received 13 October 2009; accepted 11 January 2010

Available online 8 February 2010

Communicated by C. Villani

Abstract

We establish a duality between L^p-Wasserstein control and L^q-gradient estimate in a general framework. Our result extends a known result for a heat flow on a Riemannian manifold. Especially, we can derive a Wasserstein control of a heat flow directly from the corresponding gradient estimate of the heat semigroup without using any other notion of lower curvature bound. By applying our result to a subelliptic heat flow on a Lie group, we obtain a coupling of heat distributions which carries a good control of their relative distance.

Keywords: Wasserstein distance; Gradient estimate; Subelliptic diffusion; Hamilton–Jacobi semigroup

1. Introduction

There are several ways to formulate a quantitative estimate on rate of convergence to equilibrium. By means of functional inequalities, an L^q-gradient estimate for a heat semigroup P_t

$$|\nabla P_t f|(x) \leqslant e^{-kt} P_t\big(|\nabla f|^q\big)(x)^{1/q} \tag{1.1}$$

has been known to be a very powerful tool. It implies several functional inequalities such as Poincaré inequalities (when $q = 2$) and logarithmic Sobolev inequalities (when $q = 1$), which

* Fax: +49 228 73 6716.
E-mail address: kuwada@math.ocha.ac.jp.

1 Partially supported by the JSPS fellowship for research abroad.

2 Permanent address: Graduate School of Humanities and Sciences, Ochanomizu University, Tokyo 112-8610, Japan.

0022-1236/$ – see front matter
doi:10.1016/j.jfa.2010.01.010

quantify convergence rates (see [2,4,5,21] and the references therein). As a different approach to this problem, F. Otto [30] discussed a contraction of L^p-Wasserstein distance

$$d_p^W(\mu_t, \nu_t) \leqslant \mathrm{e}^{-kt} d_p^W(\mu_0, \nu_0) \tag{1.2}$$

for two (linear or nonlinear) diffusions μ_t, ν_t of masses when $p = 2$. His heuristic observation based on the geometry of the L^2-Wasserstein space has been a source of enormous developments in the theory of optimal transport (see [37] and the references therein). To investigate a relation between these formulations makes a connection between different approaches and hence it is an interesting problem. M.-K. von Renesse and K.-T. Sturm [31] unified several formulations of this kind for linear heat equation on a complete Riemannian manifold. As a consequence of their work, (1.1) or (1.2) is shown to be equivalent to the presence of a lower Ricci curvature bound by k (it also holds for $k < 0$). But, in a more general framework, such a sort of duality has been known only when $p = 1$ and $q = \infty$, which is the weakest form for (1.1) and (1.2) both.

The main result of this paper extends the duality to that between an L^q-gradient estimate and an L^p-Wasserstein control for $p, q \in [1, \infty]$ with $p^{-1} + q^{-1} = 1$ beyond the case of a heat flow on a complete Riemannian manifold (see Theorem 2.2 for the precise statement). We should emphasize that our duality does not require any other kind of curvature conditions. An L^∞-Wasserstein control has been used in the literature as a tool to show L^1-gradient estimate in a coupling method for stochastic processes (for instance, see [38] and the references therein). In the case of heat flows in a complete Riemannian manifolds, any construction of a coupling which carries L^∞-Wasserstein control relies on lower Ricci curvature bounds. In fact, such an argument was used in von Renesse and Sturm's work. As a result, their proof employs a lower Ricci curvature bound to deduce Wasserstein controls from gradient estimates. Our result enables us to derive Wasserstein controls directly from gradient estimates. Such an implication is not known even in the case of heat flows on a Riemannian manifold. Furthermore, this is a great advantage under the lack of an appropriate notion of lower curvature bounds.

Our work is strongly motivated by recent development on gradient estimates on a Lie group endowed with a sub-Riemannian structure [5,8,12,13,22,27]. To explain a consequence of our duality, we deal with the 3-dimensional Heisenberg group here. It is the simplest example of spaces possessing a non-Riemannian sub-Riemannian structure like a flat Euclidean space in Riemannian geometry. But, unlike Euclidean spaces, some results [12,17] indicate that the "Ricci curvature" should be regarded as being unbounded from below (in a generalized sense). Nevertheless, L^q-gradient estimates hold for $q \in [1, \infty]$ with a constant $K > 1$ instead of e^{-kt} in (1.1) [5,12,13,22]. We can apply our duality to this case to obtain the corresponding L^p-Wasserstein control for any $p \in [1, \infty]$. In the theory of optimal transport on the Heisenberg group, an L^2-Wasserstein control for the heat flow would be important (cf. [18]). In probabilistic point of view, the heat flow is described by motions of a pair of the 2-dimensional Euclidean Brownian motion and the associated Lévy stochastic area. Our L^∞-Wasserstein control means the existence of a coupling of two particles so that the distance between them at time t is controlled by the initial distance almost surely. It is sometimes a complicated issue to construct a "well-behaved" coupling in the absence of curvature bounds. Especially, see [9,20] for works on a successful coupling on the Heisenberg group and its extension. Note that our formulation also fits with studying a heat semigroup under backward (super-)Ricci flow, in which case Wasserstein contractions with respect to a time-dependent distance function is shown recently [3,26].

The notion of lower Ricci curvature bound has been extended in many ways. Although our result does not need those notions, they should be related since (1.1) and (1.2) are analytic and probabilistic characterizations of a lower Ricci curvature bound respectively. Here we review two extensions and observe how these are connected with our result. In an analytic way, D. Bakry and M. Émery [6] (see also [2] and the references therein) extend the notion of lower Ricci curvature bound to Γ_2-criterion or curvature-dimension condition. In an abstract framework where it works, a Γ_2-criterion is equivalent to an L^1-gradient estimate. Note that their notion of gradient is different from ours. But, once these two notions coincide, a Γ_2-criterion becomes equivalent to L^∞-Wasserstein control with the aid of our result. In a sufficiently regular case as diffusions on a manifold, such an equivalence is well known. Our result possibly provides an extension of this equivalence. In connection with the theory of optimal transport, convexities of entropy functionals are proposed by J. Lott, C. Villani and K.-T. Sturm [24,35] as a natural extension of lower Ricci curvature bound. Under this condition, the existence of a heat flow and an L^2-Wasserstein control follow in some cases beyond Riemannian manifolds [29,33] (see [14,37] for the case on a Riemannian manifold). With the aid of Theorem 8 in [33], we can apply our duality to show an L^2-gradient estimate for the heat semigroup.

The idea of the proof of our main theorem is simple. The implication from a Wasserstein control to the corresponding gradient estimate is just a slight modification of existing arguments. The converse is based on the Kantorovich duality. If $p = 1$, the Kantorovich duality becomes the Kantorovich–Rubinstein formula and the problem becomes much simpler. In the case $p > 1$, we employ a general theory of Hamilton–Jacobi semigroup developed in [7,23] to analyze the variational formula. When $p = \infty$, we use an approximation of p by finite numbers because we are no longer able to apply the Kantorovich duality directly. Note that no semigroup property for heat semigroups is required in the proof. With keeping such a generality, our duality is sufficiently sharp in the sense that the control rate does not change when we obtain one estimate from the other, like the same e^{-kt} appears in (1.1) and (1.2) both.

The organization of this paper is as follows. In the next section, we introduce our framework and state our main theorem. We review the notion of Wasserstein distance and gradient there. Our main theorem is shown in Section 3. For the proof, we show basic properties of Wasserstein distances and summarize recent results on Hamilton–Jacobi semigroup there. In Section 4, we consider a heat flow on a sub-Riemannian manifold and apply our main theorem to these cases.

2. Framework and the main result

Let (X, d) be a complete, separable, proper, length metric space. Here, we say that d is a length metric if, for every $x, y \in X$, $d(x, y)$ equals infimum of the length of a curve joining x and y. Properness means that all closed metric balls in X of finite radii are compact. Under these assumptions, there exists a curve joining x and y whose length realizes $d(x, y)$ for each x, y (see [10], for instance). We call it minimal geodesic. Let $\tilde{d}$ be a continuous distance function on X, possibly different from d. Assume that for any $x, y \in X$, there is a minimal geodesic with respect to $\tilde{d}$ joining x and y. We call such a curve "$\tilde{d}$-minimal geodesic".

For two probability measures μ and ν on X, we denote the space of all couplings of μ and ν by $\Pi(\mu, \nu)$. That is, $\pi \in \Pi(\mu, \nu)$ means that π is a probability measure on $X \times X$ satisfying $\pi(A \times X) = \mu(A)$ and $\pi(X \times A) = \nu(A)$ for each Borel set A. For $p \in [1, \infty]$ and a measurable function $\rho : X \times X \to [0, \infty)$, we define $\rho_p^W(\mu, \nu)$ by

$$\rho_p^W(\mu, \nu) := \inf\bigl\{ \|\rho\|_{L^p(\pi)} \bigm| \pi \in \Pi(\mu, \nu) \bigr\}. \tag{2.1}$$

We are interested in the case $\rho = d$ and $\rho = \tilde{d}$. If $d_p^W(\mu, \nu) < \infty$, then there always exists a minimizer of the infimum on the right-hand side in (2.1). In addition, d_p^W satisfies all properties of distance function on the space of probability measures though it may take the value $+\infty$. The same are also true for $\tilde{d}_p^W$. These facts are well known for $p \in [1, \infty)$ and we can show it similarly even when $p = \infty$. It is sometimes reasonable to restrict d_p^W on all probability measures having finite p-th moments in order to ensure $d_p^W(\mu, \nu) < \infty$. But, in this paper, we do not adopt such a restriction. Note that, when $p < \infty$, we usually call the restriction of d_p^W the L^p-Wasserstein distance. See [36] for more details and a proof of these facts.

Let $C_b(X)$ be the space of bounded continuous functions on X equipped with the supremum norm. Let $C_L(X)$ be the collection of all Lipschitz continuous functions on X and $C_{b,L}(X) := C_b(X) \cap C_L(X)$. Note that, if we merely say "Lipschitz", it means "Lipschitz with respect to d". For Lipschitz continuity with respect to $\tilde{d}$, we use the expression "$\tilde{d}$-Lipschitz".

For a measurable function f on X and $x \in X$, we define $|\nabla_d f|(x)$ by

$$|\nabla_d f|(x) := \lim_{r \downarrow 0} \sup_{0 < d(x,y) \leqslant r} \left| \frac{f(x) - f(y)}{d(x, y)} \right|.$$

We set $\|\nabla_d f\|_\infty := \sup_{x \in X} |\nabla_d f|(x)$. Note that $\|\nabla_d f\|_\infty < \infty$ holds if and only if $f \in C_L(X)$. In addition, for $f \in C_L(X)$,

$$\|\nabla_d f\|_\infty = \sup_{x \neq y} \left| \frac{f(x) - f(y)}{d(x, y)} \right|. \tag{2.2}$$

For a pair of measurable functions f and g on X, we say that g is an upper gradient of f if, for each rectifiable curve $\gamma : [0, l] \to X$ parametrized with the arc-length, we have

$$\left| f\big(\gamma(l)\big) - f\big(\gamma(0)\big) \right| \leqslant \int_0^l g\big(\gamma(s)\big)\, ds.$$

We will use the following fact as a basic tool.

Lemma 2.1. *(See [11, Proposition 1.11], [16, Proposition 10.2].) For $f \in C_L(X)$, $|\nabla_d f|$ is an upper gradient of f.*

We also use the same notations for $\tilde{d}$. All the properties described above for $|\nabla_d f|$, including Lemma 2.1, are also true for $|\nabla_{\tilde{d}} f|$.

Set $\mathscr{P}(X)$ be the space of all probability measures on X equipped with the topology of weak convergence. Let $(P_x)_{x \in X}$ be a family of elements in $\mathscr{P}(X)$. Assume that $x \mapsto P_x$ is continuous as a map from X to $\mathscr{P}(X)$. Then $(P_x)_{x \in X}$ defines a bounded linear operator P on $C_b(X)$ by $Pf(x) := \int_X f(y) P_x(dy)$. Let P^* be the adjoint operator of P. Note that $P^*(\mathscr{P}(X)) \subset \mathscr{P}(X)$ holds.

For describing our main theorem, we state the following conditions:

Assumption 1. There exists a positive Radon measure υ on X such that

(i) (X, d, υ) enjoys the local volume doubling condition. That is, there are constants $D, R_1 > 0$ such that $\upsilon(B_{2r}(x)) \leqslant D\upsilon(B_r(x))$ holds for all $x \in X$ and $r \in (0, R_1)$.
(ii) (X, d, υ) supports a $(1, p_0)$-local Poincaré inequality for some $p_0 \geqslant 1$. That is, for every $R > 0$, there are constants $\lambda \geqslant 1$ and $C_P > 0$ such that, for any $f \in L^1_{\mathrm{loc}}(\upsilon)$ and any upper gradient g of f,

$$\int_{B_r(x)} |f - f_{x,r}| \, d\upsilon \leqslant C_P r \left\{ \int_{B_{\lambda r}(x)} g^{p_0} \, d\upsilon \right\}^{1/p_0} \tag{2.3}$$

holds for every $x \in X$ and $r \in (0, R)$, where $f_{x,r} := \upsilon(B_r(x))^{-1} \int_{B_r(x)} f \, d\upsilon$.
(iii) P_x is absolutely continuous with respect to υ for all $x \in X$; $P_x(dy) = P_x(y)\,\upsilon(dy)$. In addition, the density $P_x(y)$ is continuous with respect to x for υ-almost every $y \in X$.

Now we are in turn to state our main theorem.

Theorem 2.2. *Suppose that Assumption* 1 *holds. Then, for any $p \in [1, \infty]$, the following are equivalent*:

(i) *For all $\mu, \nu \in \mathscr{P}(X)$,*

$$d_p^W\big(P^*\mu, P^*\nu\big) \leqslant \tilde{d}_p^W(\mu, \nu). \tag{C_p}$$

(ii) *When $p > 1$, for all $f \in C_{b,L}(X)$ and $x \in X$,*

$$|\nabla_{\tilde{d}} Pf|(x) \leqslant P\big(|\nabla_d f|^q\big)(x)^{1/q}, \tag{G_q}$$

where q is the Hölder conjugate of p; $1/p + 1/q = 1$. When $p = 1$, for all $f \in C_{b,L}(X)$,

$$\|\nabla_{\tilde{d}} Pf\|_\infty \leqslant \|\nabla_d f\|_\infty. \tag{G_∞}$$

Remark 2.3. We give several remarks on Assumption 1 and Theorem 2.2.

(i) If Assumption 1(i) holds, then Assumption 1(ii) follows once we obtain (2.3) with $p_0 = 1$ for *some* $R > 0$ by a well-known argument. See [32, Lemma 5.3.1], for instance. The same is true for a $(2, 2)$-Poincaré inequality, which yield a $(1, 2)$-Poincaré inequality.
(ii) It is shown in [11] that, under Assumption 1(i)–(ii), $|\nabla_d f|$ coincides with an L^{p_0}-minimal generalized upper gradient g_f for those f for which g_f is well defined. This fact itself is not used in this article. But, it will be helpful when we apply our main theorem to more concrete problems. In fact, the notion of minimal generalized upper gradients is regarded as a sort of weak derivative in the theory of Sobolev spaces. We can identify these two notions on Euclidean spaces or Riemannian manifolds.
(iii) Assumption 1 is used only when we show the implication $(G_q) \Rightarrow (C_p)$ for $p \in (1, \infty]$. Thus the rest holds true without Assumption 1. We need Assumption 1(i)–(ii) only for employing a property of Hamilton–Jacobi semigroups. To make these facts clear, in the rest of this paper, we will mention Assumption 1 when we require it.

(iv) The duality between (1.1) and (1.2) is resumed by choosing $P = P_t$ and $\tilde{d} = \mathrm{e}^{-kt} d$. The case $\tilde{d}$ is essentially different from d naturally occurs if we consider a heat flow under a backward (super-)Ricci flow (see [3,26]).

(v) Obviously (G_q) implies $(G_{q'})$ for $q, q' \in [1,\infty]$ with $q < q'$ by the Hölder inequality. The dual implication $(C_p) \Rightarrow (C_{p'})$ for $p, p' \in [1,\infty]$ with $p > p'$ also holds true without using the equivalence in Theorem 2.2 (see Corollary 3.4 below). For a heat flow on a Riemannian manifold (i.e. $P = P_t$ and $\tilde{d} = \mathrm{e}^{-kt} d$), if (C_p) or (G_q) holds for *some* $p \in [1,\infty]$, then (C_p) and (G_q) hold for *any* $p \in [1,\infty]$. At this moment, it is not clear that what condition guarantees such an "L^p-independence".

3. Proof of Theorem 2.2

We begin with showing the implication $(C_p) \Rightarrow (G_q)$.

Proposition 3.1. *Suppose (C_p) for $p \in [1,\infty]$. Then (G_q) holds for $q \in [1,\infty]$ with $p^{-1} + q^{-1} = 1$.*

Proof. For $x, y \in X$, take $\pi_{xy} \in \Pi(P_x, P_y)$ such that $\|d\|_{L^p(\pi_{xy})} = d_p^W(P_x, P_y)$. Since $P_z = P^* \delta_z$ for $z \in X$, (C_p) yields $d_p^W(P_x, P_y) \leqslant \tilde{d}_p^W(\delta_x, \delta_y) = \tilde{d}(x, y)$. For $f \in C_{b,L}(X)$,

$$\left|Pf(x) - Pf(y)\right| = \left|\int_X f\,dP_x - \int_X f\,dP_y\right| \leqslant \int_{X\times X} \left|f(z) - f(w)\right| \pi_{xy}(dz\,dw).$$

(i) The case $p = 1$: (2.2) together with (C_1) implies

$$\int_{X\times X} \left|f(z) - f(w)\right| \pi_{xy}(dz\,dw) \leqslant \|\nabla_d f\|_\infty\, d_1^W(P_x, P_y) \leqslant \|\nabla_d f\|_\infty\, \tilde{d}(x, y).$$

Hence, by dividing the above inequalities by $\tilde{d}(x, y)$ and by taking supremum in $x \neq y$, the conclusion follows.

(ii) The case $p \in (1,\infty)$: Let us define $G_r : X \to \mathbb{R}$ by

$$G_r(z) := \sup_{w \in B_r(z)\setminus\{z\}} \left|\frac{f(z) - f(w)}{d(z, w)}\right|.$$

Set $r := \tilde{d}(x, y)^{1/(2q)}$. The Hölder inequality and the Chebyshev inequality yield

$$\begin{aligned}
&\int_{X\times X} \left|f(z) - f(w)\right| \pi_{xy}(dz\,dw) \\
&\quad = \int_{X\times X} \left|\frac{f(z) - f(w)}{d(z, w)}\right| 1_{\{0<d(z,w)\leqslant r\}}\, d(z, w)\, \pi_{xy}(dz\,dw) \\
&\qquad + \int_{X\times X} \left|f(z) - f(w)\right| 1_{\{d(z,w)>r\}}\, \pi_{xy}(dz\,dw)
\end{aligned}$$

$$\begin{aligned}
&\leqslant \left\{ \int_{X\times X} \left| \frac{f(z)-f(w)}{d(z,w)} \right|^q 1_{\{0<d(z,w)\leqslant r\}}\, \pi_{xy}(dz\,dw) \right\}^{1/q} \|d\|_{L^p(\pi_{xy})} + \frac{2\|f\|_\infty \|d\|^p_{L^p(\pi_{xy})}}{r^p} \\
&\leqslant \|G_r\|_{L^q(P_x)}\, d_p^W(P_x, P_y) + \frac{2\|f\|_\infty\, d_p^W(P_x,P_y)^p}{r^p} \\
&\leqslant \|G_r\|_{L^q(P_x)}\, \tilde{d}(x,y) + 2\|f\|_\infty\, \tilde{d}(x,y)^{1+(p-1)/2}.
\end{aligned}$$

Here the last inequality follows from (C_p). Since $\lim_{y\to x} r = 0$, $\lim_{y\to x} G_r(z) = |\nabla_d f|(z)$ holds. By virtue of $|G_r(z)| \leqslant \|\nabla_d f\|_\infty$, we can apply the dominated convergence theorem to obtain $\lim_{y\to x} \|G_r\|_{L^q(P_x)} = \||\nabla_d f|\|_{L^q(P_x)}$. Thus, by dividing the above inequalities by $\tilde{d}(x,y)$ and by tending $y \to x$, the conclusion follows.

(iii) The case $p=\infty$: (C_∞) implies $d(z,w) \leqslant \tilde{d}(x,y)$ for π_{xy}-a.e. (z,w). Hence we have

$$\int_{X\times X} \left|f(z)-f(w)\right| \pi_{xy}(dz\,dw) \leqslant \tilde{d}(x,y)\|G_{\tilde{d}(x,y)}\|_{L^1(P_x)}.$$

Thus the proof will be completed by following a similar argument as above. □

For the converse implication, first we show two auxiliary lemmas concerning to Wasserstein distances. The first one will be used to deal with L^∞-Wasserstein distance.

Lemma 3.2. *Let $\rho : X \times X \to [0,\infty)$ be a continuous function. Then*

$$\lim_{p\to\infty} \rho_p^W(\mu,\nu) = \rho_\infty^W(\mu,\nu)$$

for any $\mu,\nu \in \mathscr{P}(X)$.

Proof. Note that $\rho_p^W(\mu,\nu)$ is increasing in p by the Hölder inequality. Hence $C := \lim_{p\to\infty} \rho_p^W(\mu,\nu) \in [0,\infty]$ exists. Take $\pi_n \in \Pi(\mu,\nu)$ for $n \in \mathbb{N}$ such that $\rho_n^W(\mu,\nu) = \|\rho\|_{L^n(\pi_n)}$ holds. Since $\pi_n \in \Pi(\mu,\nu)$, $(\pi_n)_{n\in\mathbb{N}}$ is tight. Thus there exists a convergent subsequence $(\pi_{n_k})_{k\in\mathbb{N}}$ of $(\pi_n)_{n\in\mathbb{N}}$. We denote the limit of π_{n_k} by π_∞. Take $R>0$ and $n \in \mathbb{N}$ arbitrary. Since $\rho \wedge R \in C_b(X\times X)$, we have

$$\|\rho \wedge R\|_{L^n(\pi_\infty)} = \lim_{k\to\infty} \|\rho\wedge R\|_{L^n(\pi_{n_k})} \leqslant \lim_{k\to\infty} \|\rho\|_{L^{n_k}(\pi_{n_k})} = C.$$

Here the inequality follows from the Hölder inequality for sufficiently large k. Thus, as $R\to\infty$ and $n\to\infty$, we obtain $\|\rho\|_{L^\infty(\pi_\infty)} \leqslant C$. Thus the assertion holds if $\rho_\infty^W(\mu,\nu) = \infty$. When $\rho_\infty^W(\mu,\nu) < \infty$, we can take $\pi \in \Pi(\mu,\nu)$ such that $\|\rho\|_{L^\infty(\pi)} < \infty$. Then $\rho_p^W(\mu,\nu) \leqslant \|\rho\|_{L^p(\pi)} \leqslant \|\rho\|_{L^\infty(\pi)}$. Thus $C \leqslant \|\rho\|_{L^\infty(\pi)}$ holds. It yields $C \leqslant \rho_\infty^W(\mu,\nu)$ and hence the conclusion holds. □

The next one is useful to reduce the problem in a simpler case.

Lemma 3.3. *If (C_p) holds for any pair of Dirac measures, then (C_p) holds for any $\mu,\nu \in \mathscr{P}(X)$.*

Although this is probably well known for experts at least when $p \in [1, \infty)$, we give a proof for completeness.

Proof of Lemma 3.3. First we consider the case $p < \infty$. Given $\mu, \nu \in \mathscr{P}(X)$, take $\pi \in \Pi(\mu, \nu)$ so that $\|\tilde{d}\|_{L^p(\pi)} = \tilde{d}_p^W(\mu, \nu)$. We may assume $\tilde{d}_p^W(\mu, \nu) < \infty$ without loss of generality. For $x, y \in X$, take $P_{x,y} \in \Pi(P_x, P_y)$ so that $\|d\|_{L^p(P_{x,y})} = d_p^W(P_x, P_y)$. By Corollary 5.22 of [37], we can choose $\{P_{x,y}\}_{x,y \in X}$ so that the map $(x, y) \mapsto P_{x,y}$ is measurable. Define $\tilde{\pi} \in \Pi(P^*\mu, P^*\nu)$ by $\tilde{\pi}(A) := \int_{X\times X} P_{x,y}(A)\,\pi(dx\,dy)$. Then (C_p) for Dirac measures implies

$$d_p^W\big(P^*\mu, P^*\nu\big) \leqslant \|d\|_{L^p(\tilde{\pi})} = \left\{ \int_{X\times X} \|d\|_{L^p(P_{x,y})}^p \,\pi(dx\,dy) \right\}^{1/p}$$
$$\leqslant \|\tilde{d}\|_{L^p(\pi)} = \tilde{d}_p^W(\mu, \nu).$$

Thus the assertion holds. When $p = \infty$, (C_∞) for Dirac measures implies $(C_{p'})$ for Dirac measures for any $1 \leqslant p' < \infty$. Thus we obtain $(C_{p'})$ for any $\mu, \nu \in \mathscr{P}(X)$. Hence applying Lemma 3.2 for $\rho = d$ and $\rho = \tilde{d}$ yields the conclusion. $\square$

By the Hölder inequality, (C_p) for Dirac measures yields $(C_{p'})$ for Dirac measures if $p' < p$. Thus we obtain the following as a by-product of Lemma 3.3.

Corollary 3.4. *(C_p) implies $(C_{p'})$ for any $p, p' \in [1, \infty]$ with $p > p'$.*

Next we introduce the notion and some properties of Hamilton–Jacobi semigroup, which plays an essential role in the sequel. Let $L : [0, \infty) \to [0, \infty)$ be a convex superlinear function with $L(0) = 0$. Note that L is continuous and increasing. We denote the Legendre conjugate of L by $L^* : [0, \infty) \to [0, \infty)$, which is given by $L^*(z) = \sup_{w \geqslant 0}[wz - L(w)]$. For $f \in C_b(X)$ and $t > 0$, we define a function $Q_t f$ on X by

$$Q_t f(x) := \inf_{y \in X} \left[f(y) + tL\left(\frac{d(x, y)}{t}\right)\right].$$

For convenience, we write $Q_0 f := f$. We call Q_t the Hamilton–Jacobi semigroup associated with L. Several basic properties of $Q_t f$ in an abstract framework are studied in [7,23]. In [23], they assumed X to be compact and $L(s) = s^2$. In [7], they assumed $f \in C_L(X)$. Among them, the following are all we need in this paper.

Lemma 3.5. *(See [7, Theorem 2.5], [23, Theorem 2.5].)*

(i) $\inf_{y\in X} f(y) \leqslant Q_t f(x) \leqslant f(x)$. *In particular,* $Q_t f \in C_b(X)$.
(ii) $Q_t(Q_s f) = Q_{t+s} f$.
(iii) $Q_t f(x)$ *is nonincreasing in t and* $\lim_{t\downarrow 0} Q_t f(x) = f(x)$.
(iv) *Set* $u(t, x) = Q_t f(x)$. *If* $f \in C_L(X)$, *then* $u \in C_L((0, \infty) \times X)$. *Moreover,*

$$\sup_{\substack{s\neq t\\ y\neq x}} \frac{|u(t, x) - u(s, y)|}{|t - s| + d(x, y)} \leqslant \|\nabla_d f\|_\infty \vee L^*\big(\|\nabla_d f\|_\infty\big).$$

(v) *Suppose Assumption* 1(i)–(ii) *holds. Then, for $t > 0$ and υ-a.e. $x \in X$, $Q_t f$ satisfies the Hamilton–Jacobi equation associated with L^**:

$$\lim_{s \downarrow 0} \frac{Q_{t+s} f(x) - Q_t f(x)}{s} = -L^*\big(|\nabla_d Q_t f|(x)\big).$$

We do not use Lemma 3.5(ii) in the sequel. But, it explains why we call Q_t "semigroup" well. Note that Lemma 3.5(v) is shown in [7,23] for the subgradient norm instead of the gradient norm $|\nabla_d f|$. Since these two notions coincide υ-almost everywhere in this case (see [23, Remark 2.27]), Lemma 3.5(v) is still valid.

Finally, we review the Kantorovich duality (see [36, Theorem 1.3] or [37, Theorem 5.10], for example). For $\mu, \nu \in X$ and $1 \leqslant p < \infty$, the following duality holds:

$$\begin{aligned} d_p^W(\mu,\nu)^p &= \sup\left\{ \int_X g\,d\mu - \int_X f\,d\nu \;\middle|\; \begin{array}{l} f, g \in C_b(X), \\ g(y) - f(x) \leqslant d(x,y)^p \\ \text{for any } x, y \in X \end{array} \right\}, \\ &= \sup_{f \in C_b(X)} \left[\int_X f^*\,d\mu - \int_X f\,d\nu \right], \end{aligned} \tag{3.1}$$

where $f^*(y) := \inf_{x \in X}[f(x) + d(x,y)^p]$. In particular, when $p = 1$, (3.1) is written as follows:

$$d_1^W(\mu,\nu) = \sup_{\substack{f \in C_L(X) \\ \|\nabla f\|_\infty \leqslant 1}} \left[\int_M f\,d\mu - \int_M f\,d\nu \right]. \tag{3.2}$$

This is the so-called Kantorovich–Rubinstein formula (see [36, Theorem 1.14] or [37, Particular Case 5.16]).

Remark 3.6. An observation on the proof in [37] tells us that the latter supremum in (3.1) can be approximated by elements in $C_{b,L}(X)$. Actually, in that proof, there appears a sequence of pairs of functions $\phi_k, \psi_k \in C_b(X)$ approximating the former supremum in (3.1) by taking $f = \psi_k$, $g = \phi_k$. We can easily verify $\psi_k \in C_{b,L}(X)$ and that $(\psi_k)_{k \in \mathbb{N}}$ also approximates the latter supremum in (3.1). Moreover, we can assume that each element of approximating sequence has a compact support without loss of generality, thanks to the tightness of μ, ν and the properness of X.

Now we are in position to complete the proof of Theorem 2.2.

Proposition 3.7. *Suppose that Assumption* 1 *holds. Then (G_q) implies (C_p) for $p, q \in [1, \infty]$ with $p^{-1} + q^{-1} = 1$.*

Proof. By virtue of Lemma 3.3, it suffices to show (C_p) for $\mu = \delta_x$, $\nu = \delta_y$, $x \neq y$. Take a $\tilde{d}$-minimal geodesic $\gamma : [0,1] \to X$ from y to x, which is re-parametrized to have a constant speed. Here "constant speed" means $\tilde{d}(\gamma_s, \gamma_t) = |s - t|\tilde{d}(x,y)$. Note that, by (G_q), Pf is $\tilde{d}$-Lipschitz continuous if $f \in C_L(X)$.

(i) The case $p = 1$: The Kantorovich–Rubinstein formula (3.2) yields

$$d_1^W(P_x, P_y) = \sup_{\substack{f \in C_L(X) \\ \|\nabla_d f\|_\infty \leqslant 1}} \left[Pf(x) - Pf(y) \right]. \tag{3.3}$$

For $f \in C_L(X)$, we can apply Lemma 2.1 to Pf. Thus (G_∞) yields

$$\left| Pf(x) - Pf(y) \right| \leqslant \int_0^{\tilde{d}(x,y)} |\nabla_{\tilde{d}} Pf|(\gamma_s)\, ds \leqslant \|\nabla_d f\|_\infty \tilde{d}(x, y).$$

Combining this estimate with (3.3), the conclusion follows.

(ii) The case $1 < p < \infty$: Let Q_t be the Hamilton–Jacobi semigroup associated with $L(s) := p^{-1}s^p$. Note that its Legendre conjugate L^* is computed as $L^*(s) = q^{-1}s^q$. By (3.1) and Remark 3.6, we have

$$\begin{aligned} d_p^W(P_x, P_y)^p &= \sup_{f \in C_{b,L}(X)} \left[P\left(f^*\right)(x) - Pf(y) \right] \\ &= p \sup_{f \in C_{b,L}(X)} \left[PQ_1 f(x) - Pf(y) \right]. \end{aligned} \tag{3.4}$$

To obtain an integral expression of the term in the above supremum (see (3.5) below), we give some estimates. (G_q) and Lemma 3.5(i) and (iv) yield

$$|\nabla_{\tilde{d}} PQ_s f|(z) \leqslant \left\| |\nabla_d Q_s f| \right\|_{L^q(P_z)} \leqslant \|\nabla_d f\|_\infty \vee L^*\left(\|\nabla_d f\|_\infty\right)$$

for $s \geqslant 0$ and $z \in X$. Thus Lemmas 2.1 and 3.5(iv) imply

$$\begin{aligned} \left| \frac{PQ_{t+s} f(\gamma_{t+s}) - PQ_s f(\gamma_s)}{t} \right| &\leqslant \left| \frac{PQ_{t+s} f(\gamma_{t+s}) - PQ_{t+s} f(\gamma_s)}{t} \right| \\ &\quad + \left| \int_X \frac{Q_{t+s} f - Q_s f}{t}\, dP_{\gamma_s} \right| \\ &\leqslant \frac{\tilde{d}(x, y)}{t} \int_s^{t+s} |\nabla_{\tilde{d}} PQ_{t+s} f|(\gamma_u)\, du \\ &\quad + \int_X \left| \frac{Q_{t+s} f - Q_s f}{t} \right| dP_{\gamma_s} \\ &\leqslant \left(1 + \tilde{d}(x, y)\right)\left(\|\nabla_d f\|_\infty \vee L^*\left(\|\nabla_d f\|_\infty\right)\right) \end{aligned}$$

for $s \geqslant 0$. It means that $PQ_s f(\gamma_s)$ is Lipschitz continuous as a function of $s \in [0,1]$. Hence there exists a derivative $\partial_s(PQ_s f(\gamma_s))$ for a.e. $s \in [0,1]$ and we have

$$PQ_1 f(x) - Pf(y) = \int_0^1 \partial_s\bigl(PQ_s f(\gamma_s)\bigr)\,ds. \tag{3.5}$$

Let $s \in (0,1)$ be a point where $PQ_s f(\gamma_s)$ is differentiable. It implies

$$\partial_s\bigl(PQ_s f(\gamma_s)\bigr) = \lim_{t\downarrow 0} \frac{PQ_{s+t} f(\gamma_{s+t}) - PQ_s f(\gamma_s)}{t}. \tag{3.6}$$

We have

$$\begin{aligned} \frac{PQ_{s+t} f(\gamma_{s+t}) - PQ_s f(\gamma_s)}{t} &= \int_X \frac{Q_{s+t} f - Q_s f}{t}\,dP_{\gamma_{s+t}} \\ &\quad + \frac{PQ_s f(\gamma_{s+t}) - PQ_s f(\gamma_s)}{t}. \end{aligned} \tag{3.7}$$

By Lemma 2.1 together with (G_q),

$$\frac{PQ_s f(\gamma_{s+t}) - PQ_s f(\gamma_s)}{t} \leqslant \frac{\tilde{d}(x,y)}{t} \int_s^{s+t} \bigl\{P\bigl(|\nabla_d Q_s f|^q\bigr)(\gamma_u)\bigr\}^{1/q}\,du. \tag{3.8}$$

By virtue of Assumption 1(iii), the Fatou lemma together with the boundedness of $|\nabla_d Q_t f|$ implies that $(P|\nabla_d Q_s f|^q)(\gamma_u)$ is upper semi-continuous in u. Thus (3.8) yields

$$\limsup_{t\downarrow 0} \frac{PQ_s f(\gamma_{s+t}) - PQ_s f(\gamma_s)}{t} \leqslant \tilde{d}(x,y)\bigl\||\nabla_d Q_s f|\bigr\|_{L^q(P_{\gamma_s})}. \tag{3.9}$$

For the first term in (3.7), Lemma 3.5(iii) implies the integrand is nonpositive. Thanks to Assumption 1(i)–(ii), Lemma 3.5(v) is applicable to the integrand. Thus the Fatou lemma together with Assumption 1(iii) yields

$$\begin{aligned} \limsup_{t\downarrow 0} \int_X \frac{Q_{t+s} f - Q_s f}{t}\,dP_{\gamma_{s+t}} &= \limsup_{t\downarrow 0} \int_X \frac{Q_{t+s} f(z) - Q_s f(z)}{t} P_{\gamma_{s+t}}(z)\,\nu(dz) \\ &\leqslant \int_X \limsup_{t\downarrow 0} \frac{Q_{t+s} f(z) - Q_s f(z)}{t} P_{\gamma_{s+t}}(z)\,\nu(dz) \\ &= -\int_X L^*\bigl(|\nabla_d Q_s f|(z)\bigr) P_{\gamma_s}(z)\,\nu(dz). \end{aligned} \tag{3.10}$$

Combining (3.7), (3.9) and (3.10) with (3.5) and (3.6),

$$\begin{aligned} PQ_1 f(x) - Pf(y) &\leqslant \int_0^1 \left(\tilde{d}(x,y)\big\| |\nabla_d Q_s f| \big\|_{L^q(P_{\gamma_s})} - L^*\left(\big\| |\nabla_d Q_s f| \big\|_{L^q(P_{\gamma_s})}\right)\right) ds \\ &\leqslant L\big(\tilde{d}(x,y)\big), \end{aligned}$$

where the second inequality comes from the definition of L^* as the Legendre conjugate. Substituting this estimate into (3.4), we obtain the desired estimate.

(iii) The case $p = \infty$: Since (G_q) holds with $q = 1$, the Hölder inequality implies (G_q) for any $q > 1$. Thus we obtain (C_p) for any $1 \leqslant p < \infty$. Therefore, by virtue of Lemma 3.2, the conclusion follows by tending p to ∞ in (C_p). □

Remark 3.8. Our duality between L^p and L^q can be extended to a similar one between Orlicz norms. In fact, there are Hölder-type inequalities (see [1], for instance) which will be used in the implication (i) ⇒ (ii). For the converse, all properties of Hamilton–Jacobi semigroup we will use in the proof still hold in such a generality.

Remark 3.9. If (C_p) holds with $p > 1$, then we obtain the following slightly stronger version of (G_∞); for any $f \in C_{b,L}(X)$ and $x \in X$,

$$|\nabla_{\tilde{d}} Pf|(x) \leqslant \big\| |\nabla_d f| \big\|_{L^\infty(P_x)}. \tag{G'_∞}$$

As we have seen in the proof of Proposition 3.7, a weaker condition (G_∞) is sufficient to obtain (C_1). At this moment, the author does not know any example that (C_p) holds only for $p = 1$ and (G'_∞) fails.

4. Applications

In a class of sub-Riemannian manifolds, L^q-gradient estimates of a subelliptic heat semigroup is shown recently by analytic methods. In these cases, we can obtain the corresponding L^p-Wasserstein control via Theorem 2.2 though their notion of gradient looks different from ours. To explain how we deal with it, we will demonstrate a general framework of sub-Riemannian geometry generated by a family of vector fields. We refer to [16,28,34] for details.

Throughout this section, we assume X to be a finite dimensional, σ-compact, connected, smooth differentiable manifold. Consider a family of vector fields $\{X_1, \dots, X_n\}$ on X. We assume that $\{X_i(x)\}_{i=1}^n$ is linearly independent on $T_x X$ for all $x \in X$ and that $\{X_i\}_{i=1}^n$ satisfies the Hörmander condition. The latter one means that there exists a number m such that the family of vector fields generated by $\{X_i\}_{i=1}^n$ and their commutators up to the length m spans $T_x X$ for each $x \in X$. Let $\mathcal{H} \subset TX$ be the subbundle generated by $\{X_i\}_{i=1}^n$; $\mathcal{H}_x := \mathrm{Span}\{X_1(x), \dots, X_n(x)\}$. We define a metric on $\mathcal{H}$ such that $\{X_i(x)\}_{i=1}^n$ becomes an orthonormal basis of $\mathcal{H}_x$ for $x \in X$. We are interested in the case $\mathcal{H} \neq TX$. Associated with this metric, we define a function d on $X \times X$ as follows. We say a piecewise smooth curve $\gamma : [0,l] \to X$ horizontal if $\dot{\gamma}(t) \in \mathcal{H}_{\gamma(t)}$ for every t where γ is differentiable. For $x, y \in X$, we define $d(x,y)$ by

$$d(x,y) := \inf\left\{ \int_0^l \|\dot{\gamma}(t)\|_{\mathcal{H}_{\gamma(t)}}\, dt \;\middle|\; \begin{array}{l} \gamma : [0,l] \to X \text{ horizontal curve,} \\ \gamma(0) = x,\ \gamma(l) = y \end{array} \right\}.$$

By the Chow theorem, the Hörmander condition ensures that $d(x, y) < \infty$ for $x, y \in X$. As a result, the function $d : X \times X \to [0, \infty)$ becomes a distance. It is called the Carnot–Caratheodory distance. Note that the topology determined by d coincides with the original one on X. We assume that (X, d) is complete.

Let v be a Borel measure on X such that its restriction on each local coordinate has a smooth density with respect to the Lebesgue measure associated with the coordinate. Let $\Delta_{\mathcal{H}} := \sum_{i=1}^n X_i^* X_i$ be the sub-Laplacian associated with $\{X_i\}_{i=1}^n$ and v. Here X_i^* is the adjoint operator of X_i with respect to v. By the completeness of d, $\Delta_{\mathcal{H}}$ is essentially self-adjoint (see [34]). Take the self-adjoint extension of $\Delta_{\mathcal{H}}$ (also denoted by $\Delta_{\mathcal{H}}$) and consider the associated heat semigroup $P_t = \exp(t\Delta_{\mathcal{H}}/2)$. By the hypoellipticity of $\Delta_{\mathcal{H}}$, P_t has a smooth density function with respect to v. In particular, P_t becomes a Feller semigroup. We assume that P_t is conservative, i.e. $P_t 1 = 1$. For a smooth function $f : X \to \mathbb{R}$, we define the carré du champ operator $\Gamma(f) : X \to \mathbb{R}$ by $\Gamma(f)(x) := \sum_{i=1}^n |X_i f(x)|^2$.

An L^q-gradient estimate for P_t associated with Γ is formulated as follows; given $q \in [1, \infty)$, there exists $K_q(t) > 0$ for each $t > 0$ such that, for any $f \in C_c^\infty(X)$,

$$\Gamma(P_t f)(x)^{1/2} \leqslant K_q(t)\bigl\{P_t\bigl(\Gamma(f)^{q/2}\bigr)(x)\bigr\}^{1/q}, \tag{4.1}$$

where $C_c^\infty(X)$ is the set of all smooth functions $f : X \to \mathbb{R}$ with compact support. As we see in the following, (4.1) implies our gradient estimate.

Proposition 4.1. *Eq.* (4.1) *for $f \in C_c^\infty(X)$ implies* (G_q) *for $P = P_t$, $\tilde{d} = K_q(t)d$ and any $f \in C_L(X)$ with compact support.*

Proof. First we extend (4.1) for $f \in C_{b,L}(X)$. By virtue of Corollary 11.8 of [16], for $f \in C_{b,L}(X)$, the distributional derivatives $\{X_i f\}_{i=1}^n$ are represented as a bounded functions and $|\Gamma f|^{1/2} \leqslant \|\nabla_d f\|_\infty$ holds v-almost everywhere. Moreover, Theorem 11.7 of [16] implies $|\Gamma f|^{1/2} \leqslant g_f$ for any upper gradient g_f. In particular, Lemma 2.1 implies $|\Gamma f|^{1/2} \leqslant |\nabla_d f|$. Though they discussed the case that X is an open subset of a Euclidean space in [16], we can extend it to our case with the aid of a partition of unity. By a mollifier argument together with use of a partition of unity again, we can take a sequence $f_k \in C_c^\infty(X)$ such that $f_k \to f$ and $\Gamma f_k \to \Gamma f$ almost surely (cf. [16, Theorem 11.9]). Thus (4.1) holds for any $f \in C_{b,L}(X)$ with compact support.

Note that $|\Gamma f|^{1/2}$ is an upper gradient if $f \in C^\infty(X)$ (see [16, Proposition 11.6], for instance). Since $P_t f \in C^\infty(X)$ in our case, for a minimal geodesic γ joining x and y,

$$\begin{aligned} P_t f(x) - P_t f(y) &\leqslant \int_0^{d(x,y)} \bigl\{\Gamma(P_t f)\bigl(\gamma(s)\bigr)\bigr\}^{1/2}\, ds \\ &\leqslant K_q(t) \int_0^{d(x,y)} \bigl\{P_t\bigl(\Gamma(f)^{q/2}\bigr)\bigl(\gamma(s)\bigr)\bigr\}^{1/q}\, ds \\ &\leqslant K_q(t) \int_0^{d(x,y)} \bigl\{P_t\bigl(|\nabla_d f|^q\bigr)\bigl(\gamma(s)\bigr)\bigr\}^{1/q}\, ds. \end{aligned}$$

Hence the conclusion follows by dividing the above inequality by $d(x, y)$ and by letting $y \to x$. $\square$

Remark 4.2. If we suppose Assumption 1(i)–(ii) holds in Proposition 4.1, then Theorem 6.1 of [11] asserts that the minimal generalized upper gradient of f coincides with $|\nabla f|$ almost everywhere. Since the first part of the proof of Proposition 4.1 implies that $|\Gamma f|^{1/2}$ is the minimal generalized upper gradient for $f \in C_L(X)$ with compact support, the proof can be completed there in this case.

As far as the author knows, (4.1) is established in the following cases:

- The case $q = 1$ with $K_1(t) \equiv K$ for some $K > 0$ on groups of type H [13] (including the Heisenberg group of arbitrary dimension, see [5,22] also).
- The case $q > 1$ on an arbitrary Lie group [27]. Especially, $K_p(t) \equiv K_p$ for some $K_p > 0$ if it is nilpotent.
- The case $q > 1$ with $K_q(t) = K_q \mathrm{e}^{-t}$ for some $K_q > 0$ on $\mathbf{SU}(2)$ [8].

In all these cases, v is chosen to be a right-invariant Haar measure and hence the associated sub-Laplacian is of the form $\Delta_{\mathcal{H}} = \sum_{i=1}^n X_i^2$. All conditions in Assumption 1 hold in these cases. For (iii), we have already observed. By the homogeneity of the space, we can reduce the assertion in the case of a Euclidean domain (see Remark 2.3 also). Thus (i) and (ii) with $p_0 = 1$ follow from Theorems 11.19 and 11.21 of [16]. Note that (4.1) is shown on a wider class of functions than $C_c^\infty(X)$ in some cases. But it is not necessary for our purpose.

Combining Proposition 4.1 with Theorem 2.2 in these cases, we obtain (C_p) for $P = P_t$ and $\tilde{d} = K_q(t)d$. Though f is restricted to have a compact support in Proposition 4.1, it is sufficient to show (C_p) (see Remark 3.6).

The following simple examples explain a probabilistic meaning of these consequences.

Example 4.3. The 3-dimensional Heisenberg group is realized on $\mathbb{R}^3$ with the multiplication defined by

$$(x, y, z) \cdot (x', y', z') := \left(x + x', y + y', z + z' + \frac{1}{2}(xy' - yx')\right).$$

The Lebesgue measure v on $\mathbb{R}^3$ is a bi-invariant Haar measure. Let us define left-invariant vector fields X and Y by

$$X := \frac{\partial}{\partial x} - \frac{y}{2}\frac{\partial}{\partial z}, \qquad Y := \frac{\partial}{\partial y} + \frac{x}{2}\frac{\partial}{\partial z}.$$

Set $\mathcal{H} := \mathrm{Span}\{X, Y\}$. Then the diffusion process $\{\mathbf{B}_t^{\mathbf{x}}\}_{t \geqslant 0}$ associated with $\Delta_{\mathcal{H}}/2 = (X^2 + Y^2)/2$ starting at $\mathbf{x} = (x, y, z) \in \mathbb{R}^3$ is given by

$$\mathbf{B}_t^{\mathbf{x}} = \left(x + W_t^{(1)}, y + W_t^{(2)}, z + \frac{1}{2}\int_0^t \left(x + W_s^{(1)}\right) dW_s^{(2)} - \left(y + W_s^{(2)}\right) dW_s^{(1)}\right),$$

where $(W_t^{(1)}, W_t^{(2)})$ is a Brownian motion on $\mathbb{R}^2$. It means that the diffusion process associated with $\Delta_{\mathcal{H}}/2$ is given by the 2-dimensional Euclidean Brownian motion and the associated Lévy stochastic area. The corresponding heat semigroup is given by $P_t f(\mathbf{x}) = \mathbb{E}[f(\mathbf{B}_t^{\mathbf{x}})]$ for $f \in C_b(X)$. In this framework, (4.1) for $q = 1$, $P = P_t$ and $K_1(t) \equiv K$ is shown in [5,22]. Thus we obtain (C_∞). It means that, for each $t > 0$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$, there exists a coupling $(\bar{\mathbf{B}}_t^{\mathbf{x}}, \bar{\mathbf{B}}_t^{\mathbf{y}})$ of $\mathbf{B}_t^{\mathbf{x}}$ and $\mathbf{B}_t^{\mathbf{y}}$ such that

$$d\big(\bar{\mathbf{B}}_t^{\mathbf{x}}, \bar{\mathbf{B}}_t^{\mathbf{y}}\big) \leqslant K\, d(\mathbf{x}, \mathbf{y}) \tag{4.2}$$

holds almost surely. Here d is the Carnot–Caratheodory distance associated with $\mathcal{H}$. In this case, it is known that d is equivalent to the so-called Korányi distance. That is, there exist constants $C_1, C_2 > 0$ such that, for any $\mathbf{x} = (x, y, z), \mathbf{y} = (x', y', z') \in \mathbb{R}^3$,

$$\begin{aligned} C_1 d(\mathbf{x}, \mathbf{y}) &\leqslant \left\{ \big((x - x')^2 + (y - y')^2 \big)^2 + \left(z - z' + \frac{1}{2}\big(xy' - yx'\big) \right)^2 \right\}^{1/4} \\ &\leqslant C_2\, d(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Thus (4.2) is also interpreted in terms of the Korányi distance.

Remark 4.4. In Example 4.3, (C_∞) provides only a coupling of $\mathbf{B}_t^{\mathbf{x}}$ and $\mathbf{B}_t^{\mathbf{y}}$ for each fixed $t > 0$. When X is a Riemannian manifold, (C_∞) holds if and only if there exists a coupling $(\bar{\mathbf{B}}_t^{\mathbf{x}}, \bar{\mathbf{B}}_t^{\mathbf{y}})_{t \geqslant 0}$ of two Brownian motions $(\mathbf{B}_t^{\mathbf{x}})_{t \geqslant 0}$ and $(\mathbf{B}_t^{\mathbf{y}})_{t \geqslant 0}$ starting from $\mathbf{x}$ and $\mathbf{y}$ respectively such that (4.2) holds for every $t \geqslant 0$ with $K = \mathrm{e}^{-kt}$ almost surely (see [31], for instance). In Example 4.3, it is not clear whether a similar result holds or not. Actually, in Riemannian case, the fact that the constant e^{-kt} is multiplicative in $t \geqslant 0$ plays a prominent role to construct a coupling of Brownian motions from a control of their infinitesimal motions. As observed in [12], we cannot expect such a multiplicativity in the case of Example 4.3.

Example 4.5. On $\mathbb{R}^n \times \mathbb{R}^{n(n-1)/2}$, we introduce a structure of nilpotent Lie group of step 2. We index elements in $\mathbb{R}^{n(n-1)/2}$ as that in upper triangle matrices. For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n \times \mathbb{R}^{n(n-1)/2}$ written as

$$\mathbf{x} = \big((x_i)_{i=1}^n; (z_{ij})_{1 \leqslant i < j \leqslant n} \big), \qquad \mathbf{y} = \big((x_i')_{i=1}^n; (z_{ij}')_{1 \leqslant i < j \leqslant n} \big),$$

we define $\mathbf{x} \cdot \mathbf{y}$ by

$$\mathbf{x} \cdot \mathbf{y} := \left(\big(x_i + x_i'\big)_{i=1}^n; \left(z_{ij} + z_{ij}' + \frac{1}{2}\big(x_i x_j' - x_j x_i'\big) \right)_{1 \leqslant i < j \leqslant n} \right).$$

As in Example 4.3, the Lebesgue measure υ on $\mathbb{R}^n \times \mathbb{R}^{n(n-1)/2}$ becomes a bi-invariant Haar measure. Let us define left-invariant vector fields $\{X_i\}_{i=1}^n$ by

$$X_i := \frac{\partial}{\partial x_i} - \sum_{i < j \leqslant n} \frac{x_j}{2} \frac{\partial}{\partial z_{ji}} + \sum_{1 \leqslant j < i} \frac{x_j}{2} \frac{\partial}{\partial z_{ij}}.$$

Set $\mathcal{H} := \mathrm{Span}\{X_i\}_{i=1}^n$. The diffusion process $\{\mathbf{B}_t^{\mathbf{x}}\}_{t \geqslant 0}$ associated with the sub-Laplacian $\Delta_{\mathcal{H}}/2 = \sum_{i=1}^n X_i^2/2$ starting at $\mathbf{x} = ((x_i)_{i=1}^n; (z_{ij})_{1 \leqslant i<j \leqslant n}) \in \mathbb{R}^n \times \mathbb{R}^{n(n-1)/2}$ is given by

$$\mathbf{B}_t^{\mathbf{x}} = \left(\left(x_i + W_t^{(i)}\right)_{i=1}^n; \left(z_{ij} + \frac{1}{2} \int_0^t \left(x_i + W_s^{(i)}\right) dW_s^{(j)} - \left(x_j + W_s^{(j)}\right) dW_s^{(i)} \right)_{1 \leqslant i<j \leqslant n} \right).$$

We can easily verify that this group is of type H only if $n = 1$ (see Corollary 1 of [19], for example). But it is still in the framework of [27]. Thus, for each $p \in [1, \infty)$, there is a constant $K_p > 0$ such that, for any pair $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n \times \mathbb{R}^{n(n-1)/2}$, there is a coupling $(\bar{\mathbf{B}}_t^{\mathbf{x}}, \bar{\mathbf{B}}_t^{\mathbf{y}})$ of $\mathbf{B}_t^{\mathbf{x}}$ and $\mathbf{B}_t^{\mathbf{y}}$ satisfying

$$\mathbb{E}\left[d\left(\bar{\mathbf{B}}_t^{\mathbf{x}}, \bar{\mathbf{B}}_t^{\mathbf{y}}\right)^p\right]^{1/p} \leqslant K_p\, d(\mathbf{x}, \mathbf{y}). \tag{4.3}$$

Finally, we give a remark that a different kind of coupling of this process is studied by Kendall [20]. He showed the existence of a successful coupling. As mentioned there, studying a coupling of this process has a possibility of a future application to rough path theory [15,25].

References

[1] R. Adams, J. Fournier, Sobolev Spaces, 2nd edition, Pure Appl. Math. (Amst.), vol. 140, Elsevier/Academic Press, Amsterdam, 2003.

[2] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, G. Scheffer, Sur les inégalités de Sobolev logarithmiques, Panor. Synthèses, vol. 10, Société Mathématique de France, Paris, 2000.

[3] M. Arnaudon, K. Coulibaly, A. Thalmaier, Horizontal diffusion in C^1-path space, in: Séminaire de Probabilités, in: Lecture Notes in Math., 2009, in press, arXiv:0904.2762.

[4] D. Bakry, On Sobolev and logarithmic Sobolev inequalities for Markov semigroups, in: New Trends in Stochastic Analysis, Charingworth, 1994, World Sci. Publ., River Edge, NJ, 1997, pp. 43–75.

[5] D. Bakry, F. Baudoin, M. Bonnefont, D. Chafaï, On gradient bounds for the heat kernel on the Heisenberg group, J. Funct. Anal. 255 (8) (2008) 1905–1938.

[6] D. Bakry, M. Émery, Hypercontractivité de semi-groupes de diffusion, C. R. Acad. Sci. Paris Sér. I Math. 299 (15) (1984) 775–778.

[7] Z. Balogh, A. Engoulatov, L. Hunziker, O. Maasalo, Functional inequalities and Hamilton–Jacobi equations in geodesic spaces, preprint, arXiv:0906.0476.

[8] F. Baudoin, M. Bonnefont, The subelliptic heat kernel on $\mathbf{SU}(2)$: Representations, asymptotics and gradient bounds, Math. Z. 263 (3) (2009) 647–672.

[9] G. Ben Arous, M. Cranston, W. Kendall, Coupling constructions for hypoelliptic diffusions: two examples, in: Stochastic Analysis, Ithaca, NY, 1993, in: Proc. Sympos. Pure Math., vol. 57, American Mathematical Society, Providence, RI, 1995, pp. 193–212.

[10] D. Burago, Y. Burago, S. Ivanov, A Course in Metric Geometry, Grad. Stud. Math., vol. 33, American Mathematical Society, Providence, RI, 2001.

[11] J. Cheeger, Differentiability of Lipschitz functions on metric measure spaces, Geom. Funct. Anal. 9 (3) (1999) 428–517.

[12] B. Driver, T. Melcher, Hypoelliptic heat kernel inequalities on the Heisenberg group, J. Funct. Anal. 221 (5) (2005) 340–365.

[13] N. Eldredge, Gradient estimates for the subelliptic heat kernel on H-type groups, J. Funct. Anal. 258 (2) (2010) 504–533.

[14] M. Erbar, The heat equation on manifolds as a gradient flow in the Wasserstein space, Ann. Inst. H. Poincaré, in press.

[15] P. Friz, N. Victoir, Multidimensional Stochastic Processes as Rough Paths: Theory and Applications, Cambridge University Press, in press.

[16] P. Hajłasz, P. Koskela, Sobolev met Poincaré, Mem. Amer. Math. Soc. 145 (688) (2000) 1–101.

[17] N. Juillet, Geometric inequalities and generalized Ricci bound on the Heisenberg group, Int. Math. Res. Not. IMRN 2009 (13) (2009) 2347–2373.
[18] N. Juillet, Diffusion by optimal transport on Heisenberg groups, preprint.
[19] A. Kaplan, Fundamental solutions for a class of hypoelliptic PDE generated by composition of quadratic forms, Trans. Amer. Math. Soc. 258 (1) (1980) 147–153.
[20] W.S. Kendall, Coupling all the Lévy stochastic areas of multidimensional Brownian motion, Ann. Probab. 35 (3) (2007) 935–953.
[21] M. Ledoux, The geometry of Markov diffusion generators, Ann. Fac. Sci. Toulouse Math. (6) 9 (2) (2000) 305–366.
[22] H.-Q. Li, Estimation optimale du gradient du semi-groupe de la chaleur sur le groupe de Heisenberg, J. Funct. Anal. 236 (2) (2006) 369–394.
[23] J. Lott, C. Villani, Hamilton–Jacobi semigroup on length spaces and applications, J. Math. Pures Appl. (9) 88 (3) (2007) 219–229.
[24] J. Lott, C. Villani, Ricci curvature for metric-measure spaces via optimal transport, Ann. of Math. 169 (3) (2009) 903–991.
[25] T. Lyons, Z. Qian, System Control and Rough Paths, Oxford University Press, 2002.
[26] R. McCann, P. Topping, Ricci flow, entropy and optimal transportation, Amer. J. Math., in press.
[27] T. Melcher, Hypoelliptic heat kernel inequalities on Lie groups, Stochastic Process. Appl. 118 (3) (2008) 368–388.
[28] R. Montgomery, A Tour of Subriemannian Geometries, Their Geodesics and Applications, Math. Surveys Monogr., vol. 91, American Mathematical Society, Providence, RI, 2002.
[29] S.-i. Ohta, Gradient flows on Wasserstein spaces over compact Alexandrov spaces, Amer. J. Math. 131 (2) (2009) 475–516.
[30] F. Otto, The geometry of dissipative evolution equations: the porous medium equation, Comm. Partial Differential Equations 26 (1–2) (2001) 101–174.
[31] M.-K. von Renesse, K.-T. Sturm, Transport inequalities, gradient estimates, entropy and Ricci curvature, Comm. Pure Appl. Math. 58 (7) (2005) 923–940.
[32] L. Saloff-Coste, Aspects of Sobolev-type Inequalities, London Math. Soc. Lecture Note Ser., vol. 289, Cambridge University Press, Cambridge, 2002.
[33] G. Savaré, Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds, C. R. Acad. Sci. Paris 345 (3) (2007) 151–154.
[34] R.S. Strichartz, Sub-Riemannian geometry, J. Differential Geom. 24 (2) (1986) 221–263.
[35] K.-T. Sturm, On the geometry of metric measure spaces. I, Acta Math. 196 (1) (2006) 65–131.
[36] C. Villani, Topics in Optimal Transportations, Grad. Stud. Math., vol. 58, American Mathematical Society, Providence, RI, 2003.
[37] C. Villani, Optimal Transport, Old and New, Grundlehren Math. Wiss., vol. 338, Springer-Verlag, 2008.
[38] F.-Y. Wang, On estimation of the logarithmic Sobolev constant and gradient estimates of heat semigroups, Probab. Theory Related Fields 108 (1) (1997) 87–101.